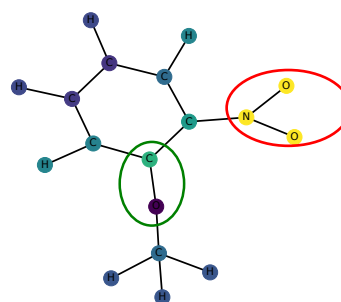




## Explaining and Visualizing Graph Neural Networks

Graph Neural Networks (GNNs) are bringing the power of deep neural networks to the domain of graph-structured problems. This allows prediction and classification tasks on a variety of interesting problems. For these reasons, GNNs are now a hot topic for further research and new applications. We want to understand better what GNNs learn and how they make their decisions. For the graph on the left, one method explains the prediction with an  $NO_2$  group and the  $CO$  connection nearby.



Generally, GNNs share similarities to Convolutional Neural Networks. We can see CNNs as operating on special kinds of graphs, regular grid graphs. In CNNs, pixels can be seen as nodes that are connected to all adjacent pixels. However, applying the explanation techniques from CNNs to GNNs does not bring the desired results. The operations used in CNNs do tiny modifications, but even these modifications can have drastic effects on graphs. For example, removing a single node can create a disconnected graph, which was never seen during training. This entirely new graphs can confuse the model and give misleading explanations. We did previous studies on explanation techniques that avoid this problem that we want to continue and come up with new angles at this problem.

**Requirements:** Knowledge in deep learning or a strong interest is advantageous for this project. We also recommend programming experience since this project will involve a fair amount of programming. We will have weekly meetings to discuss the intermediate progress, think together about future ideas, and tackle open questions.

**Interested? Please contact us for more details!**

- Lukas Faber: [lfaber@ethz.ch](mailto:lfaber@ethz.ch), ETZ G93